



THE SECOND GRAND CHALLENGE ON NEURAL NETWORK-BASED VIDEO CODING WITH ISCAS 2023



ABSTRACT

Recently, there is an increasing interest in the neural network-based video coding, in both academia and standardization. To foster the research in this emerging field, we organized a grand challenge on neural network-based video coding with ISCAS 2022 and received many responses from various universities and companies. To continue this activity and track the progress year by year, we propose this challenge, i.e., the 2nd grand challenge on neural network-based video coding with ISCAS 2023. Like the 1st grand challenge, in this GC, different neural network-based coding schemes will be evaluated according to their coding efficiency and innovations in methodologies. Two tracks will be evaluated, including the hybrid and end-to-end solutions. In the hybrid solutions, deep network-based coding tools shall be used with traditional video coding schemes. In the end-to-end solutions, the whole video codec system shall be built primarily upon deep networks. Participants shall express their interest to participate in this Grand Challenge by sending an email to the organizer Dr. Yue Li (yue.li@bytedance.com) and are invited to submit their proposals as ISCAS papers. The papers will be regularly reviewed and, if accepted, must be presented at ISCAS 2023. The submission instructions for Grand Challenge papers will be communicated by the organizers.

RATIONALE

In recent years, deep learning-based image/video coding schemes have achieved remarkable progress. As two representative approaches, hybrid solutions and end-to-end solutions have both been investigated extensively. The hybrid solution adopts deep network-based coding tools to enhance traditional video coding schemes while the end-to-end solution builds the whole compression scheme based on deep networks. In spite of the great advancement of these solutions, there are still numerous challenges remaining to be addressed, e.g.: **1)** how to harmonize a deep coding tool with a hybrid video codec, e.g. how to take compression into consideration when developing a deep tool for pre-processing; **2)** how to exploit long-term temporal dependency in an end-to-end framework for video coding; **3)** how to leverage automated machine learning-based network architectures optimization for higher coding efficiency; **4)** how to perform efficient bit allocation with deep learning frameworks; **5)** how to achieve the global minimum in rate-distortion trade-offs, e.g. to take the impact of the current step on later frames into account, possibly by using reinforcement learning; and **6)** how to achieve better complexity-performance trade-offs. In view of these challenges, several activities towards improving deep-learning-based image/video coding schemes have been initiated. For example, there are a special section on "Learning-based Image and Video Compression" in TCSVT, July 2020, a special section on "Optimized Image/Video Coding Based on Deep Learning" in OJCAS, Dec. 2021, and the "Challenge on Learned Image Compression (CLIC)" in CVPR, which has been organized annually since 2018. Meanwhile, JPEG started the JPEG-AI project targeting at a neural network-based image compression standard; and JVET also started to explore neural network-based video coding technologies for the potential next generation video coding standard. In hopes of encouraging more innovative contributions towards resolving the aforementioned challenges in the ISCAS community, we propose this grand challenge.

Training Data Set

It is recommended to use the following training data.

1. UVG dataset, <http://ultravideo.cs.tut.fi/>
2. CDVL dataset, <https://cdvl.org/>

Additional training data are also allowed to be used given that they are described in the submitted document.

Test Specifications

In the test, the proposals will be evaluated on multiple YUV 4:2:0 test sequences in the resolution of 1920x1080. There is no constraint on the reference structure. Note that the neural network must be used in the decoding process.

Evaluation Criteria

The test sequences will be released according to the timeline in Table 1 and the results will be evaluated with the following criteria:

1. The decoded sequences will be evaluated in 4:2:0 color format.
2. PSNR $(6 \cdot \text{PSNRY} + \text{PSNRU} + \text{PSNRV})/8$ will be used to evaluate the distortion of the decoded pictures.
3. Average Bjøntegaard delta PSNR (BD-PSNR) calculated using [1] for all test sequences will be gathered to compare the coding efficiency.
4. An anchor of HM 16.22 [2] coded with QPs = {22, 27, 32, 37} under random access configuration defined in the HM common test conditions [3] will be provided. The released anchor data will include the bit-rates corresponding to the four QPs for each sequence. It is required that the proposed method should generate four bit-streams for each sequence, targeting the anchor bit-rates corresponding to the four QPs. Additional constraints are listed as follows:
 - a. For each sequence, the bit-rate difference of the lowest bit-rate points between the anchor and the test shall be less than 20%; the bit-rate difference of the highest bit-rate points between the anchor and the test shall be less than 20%.
 - b. Only one single decoder shall be utilized to decode all the bitstreams.
 - c. The intra period in the proposed submission shall be no larger than that used by the anchor in generating the validation and test sequences.

Proposed documents

A docker container with the executable scheme must be submitted for results generation and cross-check. Each participant is invited to submit an ISCAS paper, which must describe the following items in detail.

- The methodology;
- The training data set;
- Detailed rate-distortion data (Comparison with the provided anchor is encouraged).

Important Dates

Aug. 15, 2022	The organizers release the validation set as well as the corresponding test information (e.g., frame rates and intra periods) and template for performance reporting (with rate-distortion points for the validation set)
Oct. 24, 2022	Deadline of paper submission (to be aligned with Special Sessions in case of extension) for participants
Nov. 08, 2022	Participants upload docker container wherein only one single decoder shall be utilized for the decoding of all the bitstreams
Nov. 10, 2022	The organizers release the test sequences (including frame rate, corresponding rate-distortion points, etc.)
Dec. 01, 2022	Participants upload compressed bitstreams and decoded YUV files
Dec. 14, 2022	Deadline of fact sheets submission for participants
Dec. 19, 2022	Paper acceptance notification
Feb. 04, 2023	Camera-ready paper submission deadline
TBA	Paper presentation at ISCAS 2023
TBA	Awards announcement (at the ISCAS 2023 banquet)



Awards

ByteDance will sponsor the awards of this grand challenge. Three categories of awards are expected to be presented. Two top-performance awards will be granted according to the performance, for the hybrid track and the end-to-end track, respectively. In addition, to foster the innovation, a top-creativity award will be given to the most inspiring scheme recommended by a committee group, and it is only applicable to participants whose papers are accepted by ISCAS 2023. The winner of each award (if any) will receive a \$5000 USD prize.

References

- [1] G. Bjøntegaard, "Calculation of average PSNR differences between RD-Curves," ITUT SG16/Q6, Doc. VCEG-M33, Austin, Apr. 2001.
- [2] <https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tree/HM-16.22>
- [3] Common Test Conditions and Software Reference Configurations for HM (JCTVC-L1100)

Organizer Biographies

Li Zhang received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2009. From 2009 to 2011, she held a post-doctoral position at the Institute of Digital Media, Peking University, Beijing. From 2011 to 2018, she was a Senior Staff Engineer at Qualcomm, Inc., San Diego, CA, USA. She is currently the Head of Multimedia Lab, Bytedance Inc., San Diego, CA, USA. Her research interests include 2D/3D image/video coding, video processing, and transmission. She was a Software Coordinator for Audio and Video Coding Standard (AVS) and the 3D extensions of High Efficiency Video Coding (HEVC). She has authored 600+ standardization contributions, 300+ granted US patents, 100+ technical articles in related book chapters, journals, and proceedings in image/video coding and video processing. She has been an active contributor to the Versatile Video Coding (VVC), Advanced AVS, the IEEE 1857, 3D Video (3DV) coding extensions of H.264/AVC and HEVC, and HEVC screen content coding extensions. During the development of those video coding standards, she co-chaired several ad hoc groups and core experiments. She has been appointed as an Editor of AVS, the Main Editor of the Software Test Model for 3DV Standards. She organized/co-chaired multiple special sessions and grand challenges at various conferences.

Jizheng Xu received the Ph.D. degree in electrical engineering from Shanghai Jiaotong University, China in 2011. He joined Microsoft Research Asia in 2003 and served as a Research Manager and joined ByteDance multimedia lab as a Research Scientist in 2018. He has authored and co-authored over 140 refereed conference and journal refereed papers. His research interests include image and visual signal representation, image/video compression and communication, computer vision and deep learning. He has been an active contributor to ISO/MPEG and ITU-T video coding standards, including H.264/AVC, H.265/HEVC and VVC/H.266. He initiated the screen content coding in H.265/HEVC and was a major technical contributor. He chaired and co-chaired the ad-hoc group of exploration on wavelet video coding in MPEG, and various technical ad-hoc groups in JCT-VC, e.g., on screen content coding, on parsing robustness, on lossless coding. He was an Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology from 2018 to 2020. He served as a Guest Editor for the special issue on Screen Content Video Coding and Applications of the IEEE Journal on Emerging and Selected Topics in Circuits and Systems in 2016. He co-organized and co-chaired special sessions on scalable video coding, directional transform, high quality video coding at various conferences. He is an Associate Editor for the IEEE Transactions on Multimedia since 2022.

Kai Zhang received the B.S. degree in computer science from Nankai University, Tianjin, China, in 2004. In 2011, he received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. From 2011 to 2012, he worked as a researcher in Tencent Inc. Beijing, China. From 2012 to 2016, he worked as a team manager in Mediatek Inc. Beijing, China, leading a research team to propose novel technologies to emerging video coding standards. From 2016 to 2018, he worked in Qualcomm Inc. San Diego, CA, still focusing on video coding standardization. Now, he is doing research work on video coding and leading the standardization team in Bytedance Inc. San Diego, CA. Dr. Zhang's research interests include video/image compression, coding, processing and communication, especially video coding standardization. In 2006, he proposed his first proposal to JVT. From then, he has contributed more than 500 proposals to JVT, VCEG, JCT-VC, JCT-3V, JVET and AVS, covering many important aspects of major standards such as H.264/AVC, HEVC, 3D-HEVC, VVC and AVS-1,2,3. Dr. Zhang has 500+ granted or pending U.S. patents applications.

Most of these patents are essential to popular video coding standards. During the development of VVC, Dr. Zhang co-chaired several core experiments and branch of groups. Currently, Dr. Zhang serves as a coordinator of the reference software known as ECM in JVET, to explore video coding technologies beyond VVC. Dr. Zhang has co-authored 40+ journal or conference papers. He also serves as a reviewer for 70+ papers of wellknown journals and conferences.

Yue Li received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2014 and 2019, respectively. He is currently a Research Scientist with Bytedance Multimedia Lab, San Diego, CA, USA. His research interests include image/video coding and processing. He has authored 20+ neural network-based standardization contributions to the JVET. Currently, Dr. Li serves as a coordinator of the common reference software on neural network-based video coding in JVET. He has authored/co-authored 20+ papers on well-known journals/conferences such as T-IP, T-CSVT, TOMM, CSUR, ICIP, ICME, DCC, etc. He also serves as a reviewer for those journals/conferences.

Sponsor



Technical Sponsor

